The EMBO Journal vol.13 no.19 pp.4704–4714, 1994

DNA bending creates favored sites for retroviral integration: an explanation for preferred insertion sites in nucleosomes

Hans-Peter Müller[1] and Harold E.Varmus[2]

Department of Microbiology and Immunology, University of California, San Francisco, CA 94143-0502, USA

[1]Present address: Institut für Molekularbiologie II der Universität Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland

[2]Present address: Office of the Director, NIH, Building 1, Room 126, 9000 Rockville Pike, Bethesda, MD 20892, USA

Communicated by H.E.Varmus

The choice of retroviral integration sites is strongly influenced by chromatin: integration *in vitro* occurs more efficiently into nucleosomal DNA than into naked DNA, and a characteristic pattern of preferred insertion sites with a 10 bp periodicity is observed at the outer face of the nucleosomal DNA. At least three features of nucleosomal DNA could be responsible for the creation of these favored sites: the presence of histones, attachment of the DNA to a protein surface, and DNA bending. To test each of these possibilities, we studied integration *in vitro* with human immunodeficiency virus and murine leukemia virus integrases into four model targets that mimic features of nucleosomal DNA: (i) catabolite activator protein–DNA complexes; (ii) *lac* repressor–operator complexes; (iii) *lac* repressor-induced loops; and (iv) intrinsically bent A-tract DNA. We found that bending of the target DNA can create favored integration sites at the outer face of the helix, irrespective of whether the bent DNA is attached to a protein surface. Our findings offer an explanation for the preferred usage of nucleosomes as integration targets. In addition, they suggest that bending of the target DNA might be an intrinsic feature of the integration reaction.

Key words: CAP/DNA bending/*lac* repressor/nucleosome/retroviral integration

Introduction

During retroviral infection, a DNA copy of the viral RNA genome is inserted into the host cell genome in a recombination process termed integration (Varmus and Brown, 1989; Coffin, 1990). The components known to be necessary for the catalysis of integration are provided by the virus: integrase (IN), a protein encoded at the 3′-end of the *pol* reading frame, and *att* sites, short inverted repeats located at the ends of the viral DNA. IN recognizes the *att* sites in a sequence-specific fashion and prepares them for integration by removing two nucleotides from the 3′-ends (3′-processing). Subsequently, IN mediates the insertion of the linear viral DNA into the host genome by a nucleophilic attack of the 3′-ends of the viral DNA at phosphates in the target DNA (strand transfer; for reviews on retroviral integration see Goff, 1992; Mizuuchi, 1992; Whitcomb and Hughes, 1992; Vink and Plasterk, 1993).

The choice of where integration into the host cell genome occurs could be influenced by target DNA sequence and conformation, as well as host proteins associated with the cellular genome. To learn how integration is affected by one aspect of nuclear organization, namely chromatin structure, Pryciak and Varmus (1992) studied integration site selection *in vitro* using simple chromatin targets such as minichromosomes and reconstituted mononucleosomes. Murine leukemia virus (MLV) nucleoprotein complexes isolated from infected cells, as well as purified MLV IN and human immunodeficiency virus (HIV) IN, were used as integration activities. The distribution of integration sites was measured using a PCR-based assay that allows the positions and frequencies of a very large number of integration events to be scored. The following observations were made: (i) integration into naked DNA is non-random, suggesting some bias for sequence; (ii) integration into nucleosomes occurs more efficiently than into nucleosome-free regions; (iii) preferred sites on nucleosomal DNA occur with a 10 bp periodicity; and (iv) comparison of integration sites (obtained with MLV nucleoprotein complexes) and DNase I hypersensitive sites on a rotationally phased mononucleosome revealed that preferred integration events occur where the major groove is exposed. Remarkably, observations (ii) and (iii) were made with purified INs as well as nucleoprotein complexes, indicating that the additional viral proteins found in nucleoprotein complexes (Bowerman *et al.*, 1989) and the ability of nucleoprotein complexes to perform concerted two-end strand transfer reactions (Fujiwara and Mizuuchi, 1988; Brown *et al.*, 1989) are not essential for the stimulatory effects observed with nucleosomal DNA. Purified IN proteins perform primarily 'single-end' strand transfer reactions (Bushman *et al.*, 1990; Craigie *et al.*, 1990; Katz *et al.*, 1990; Fitzgerald *et al.*, 1992).

We have previously suggested that at least three properties of nucleosomes might alter the selection of integration sites on nucleosomal DNA: (i) the presence of histones; (ii) the attachment of DNA to a protein surface; and (iii) DNA bending (Müller *et al.*, 1993). We now present the results of *in vitro* experiments that address these possibilities by using purified MLV IN and HIV IN and model targets that mimic some, but not all, features of nucleosomal DNA. Our results show that bending of the target DNA by various means allows integration at the outer face of the DNA helix to occur with similar or even higher efficiency than on the unbent target DNA. In contrast, integration at the inner face of the bend is never enhanced, but often reduced or blocked. The effects are observed irrespective of whether the bent DNA is attached to a protein surface. DNA bending around nucleosomal

cores is likely to govern integration site selection on nucleosomes in a similar fashion. Furthermore, the stimulatory effect of DNA bending on integration suggests that DNA bending is an intrinsic step in the integration reaction.

Results

Model targets that mimic features of nucleosomal DNA

To understand the features of nucleosomal DNA responsible for the pattern of preferred integration sites previously observed (Pryciak and Varmus, 1992), we studied integration by purified MLV IN and HIV IN into four model targets, each of which recapitulates at least one feature of nucleosomal DNA (Table I). These models should help determine which properties of nucleosomes contribute to integration site selection. In catabolite activator protein (CAP)–DNA complexes, DNA is bent over a protein surface other than histones. In contrast, in *lac* repressor–operator complexes, DNA is attached to a protein surface but is not bent. Looping of DNA by *lac* repressor provides DNA that is bent but not attached to a protein surface. Finally, A-tract DNA is intrinsically bent due to sequence and requires no associated protein.

CAP–DNA complexes as integration targets

Sequence-specific binding of CAP creates an overall bend of ~90° in the DNA at the binding site (Schultz *et al.*, 1991; for review on CAP see Crothers and Steitz, 1992). Consequently, CAP induces two DNase I hypersensitive sites that are 10 bp apart and located within the binding site (Schmitz, 1981; Spassky *et al.*, 1984). In these ways CAP–DNA complexes recapitulate two of the three features of nucleosomal DNA as indicated in Table I.

The CAP binding site of the *Escherichia coli lac* operon was used as a target for MLV IN and HIV IN in the presence or absence of CAP binding. The integration recombinants were amplified by PCR and the products separated on a denaturing gel to survey integration site distribution (Prycak and Varmus, 1992). CAP binding affects integration by MLV IN and HIV IN in similar ways (Figure 1A and 1B). On one hand, CAP blocks sites for integration, producing an integration footprint (Figure 1A: compare lanes 2 and 3, 4 and 5, 6 and 7, and 8 and 9; footprints indicated by black bars). On the other hand, at three clusters of sites within the footprint, CAP binding allows integration to occur at least as efficiently as on naked DNA (Figure 1B, black arrowheads), and some of these sites are used more efficiently than in naked DNA (Figure 1A and B; black arrowheads with position numbers). The enhancement of individual sites on CAP–DNA complexes versus naked DNA ranges from 2- to 4-fold as measured by phosphorimaging (not shown). The clusters of integration sites within the footprint of CAP-bound DNA are separated by ~10 bp on each strand, suggesting some preferred feature is repeated in each turn of the helix. As indicated in Figure 1B, neighboring clusters on opposite strands are staggered relative to each other similar to the 4-6 bp 5′ stagger normally seen during retroviral integration *in vivo* (for a review see Goff, 1992). This is noteworthy because purified MLV IN and HIV IN confer primarily 'single-end' integration events *in vitro* (Bushman *et al.*, 1990; Craigie *et al.*, 1990), and it may suggest that integration by purified MLV IN and HIV IN occurs where the major groove is exposed, as observed with MLV nucleoprotein complexes (Pryciak and Varmus, 1992). However, we cannot exclude the formal possibility that the ~5 bp 5′ stagger between integration sites on opposite strands at least partly reflects accessibility to sites on the target DNA. Also, since insertions catalyzed by purified IN are predominantly 'single-end' attacks on one strand of the target DNA, integration sites cannot be unambiguously assigned as pairs.

Incorporation of the integration sites on CAP-bound DNA into the crystallographic structure model of the CAP–DNA complex (Schultz *et al.*, 1991) revealed the following (Figure 1C). (i) The favored integration sites within the footprint region occur on the outer, exposed face of the DNA helix, away from the CAP protein surface. (ii) The sites that are not used for integration when CAP is bound are generally located at the inner face of the helix, closer to the CAP protein surface, indicating that steric exposure is an important prerequisite for a site to be used by IN. Hence, the footprint observed in Figure 1A is probably due to steric interference between CAP and IN. (iii) Those sites that are more efficiently used on CAP-bound DNA than on naked DNA are not always the most exposed ones, but instead some of them are shifted slightly towards the CAP protein surface (Figure 1C, panel 6; see, for example, the two sites on the left). This may imply that some other structural feature of the DNA—for example, the bend geometry—is more important then steric exposure.

Overall, the distribution of preferred integration sites on CAP-bound DNA is reminiscent of what has been previously observed on nucleosomal DNA (Pryciak and Varmus, 1992), namely, inhibition at sites in contact with the protein core, while sites on the exposed face of the helix remain accessible or become even more reactive. Therefore, the presence of histones is not absolutely

Table I. Schematic view of model targets for retroviral integration that mimic features of nucleosomal DNA

	 Nucleosome	 CAP–DNA complex	 *lac* repressor–DNA complex	 *lac* repressor-induced loop	 A-tract DNA
Histones present	+	–	–	–	–
DNA on surface	+	+	+	–	–
DNA bent	+	+	–	+	+

Symbols: +, nucleosomal feature present; –, nucleosomal feature absent.

required for any of these effects, although the degree of increased reactivity may not be as high in the CAP–DNA target as in nucleosomal targets (see Discussion).

Lac repressor-induced loops as integration targets

The experiments presented in the previous section suggest that DNA bending, the presence of a protein surface, or both, contribute to the observed distribution of integration sites. To distinguish between these two features, we next tested looped DNA molecules produced by *lac* repressor as integration targets.

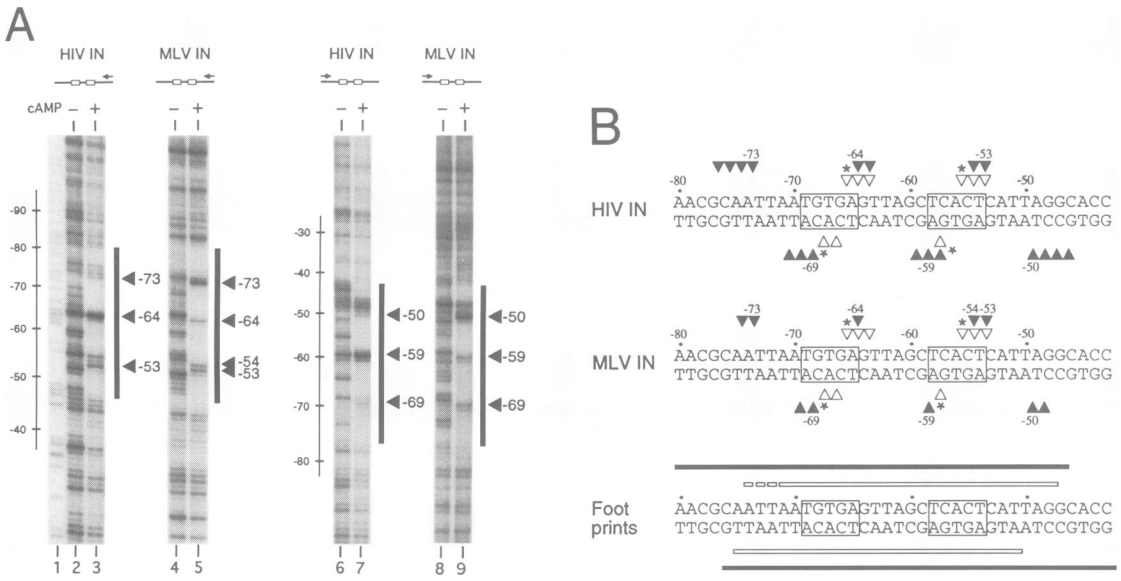
When presented with a DNA molecule containing two operators separated by an even number of half helix turns, the tetrameric *lac* repressor can bind simultaneously to both operators, thereby looping out the intervening DNA (Krämer *et al.*, 1987; for review see Gralla, 1992). Under appropriate conditions, almost 100% of the molecules can be looped as measured by a gel shift assay (Krämer *et al.*, 1987; Müller *et al.*, 1993). This system is of particular value to determine whether DNA bending affects integration site selection, because (i) the same DNA sequence can be tested in both straight and bent configurations; (ii) the degree of bending can be varied by changing the size of the spacer between the two *lac* operators; (iii) the looped region of the DNA is not attached to a protein; and (iv) the DNase I digestion pattern is sensitive to looping in this system (Krämer *et al.*, 1987; see also below) and can therefore be used as a positive control.

In a first experiment, MLV IN and HIV IN were tested with the construct *lac*-52-*lac*, in which the two operators are separated by 52 bp (measured from the center of one operator to the center of the other). When *lac* repressor is bound, integration into the operator region is blocked, producing integration footprints (Figure 2A: compare lanes 3 and 4 with 1 and 2, and lanes 7 and 8 with 5 and 6). In contrast to the CAP–DNA complexes, in which some sites within the footprint remain accessible or become even more reactive (Figure 1A), integration is inhibited at all sites within the region bound by *lac* repressor, similar to the situation previously reported with the yeast α2 repressor (Pryciak and Varmus, 1992).

Importantly, the distribution of integration sites in the spacer DNA between the two *lac* operators is dependent upon whether the spacer is straight or looped (Figure 2A). In contrast, the integration pattern in the areas outside of the looped region adjacent to the operators is independent of *lac* repressor binding and looping (Figure 2A, compare lanes above the upper footprint and below the lower footprint). This implies that the effect of *lac* repressor on integration within the DNA loop is due to bending.

In the looped DNA, integration is reduced at some sites (⊢) and enhanced at other sites (black arrowheads). The positions of enhanced integration are separated by roughly 10 bp. As measured with a PhosphorImager, integration at some of these sites is enhanced between 2- and 4-fold (not shown). The integration sites preferred by MLV IN and HIV IN occur at similar positions (Figure 2C). The distribution of these sites relative to the DNase I hypersensitive sites (experiment in Figure 2B) is similar to that observed with CAP-bound DNA (compare Figures 2C and 1B), suggesting that the preferred integration sites are located again at the outer face of the bend. Note that not every potential site on the looped DNA is preferred by both MLV IN and HIV IN. With HIV IN, for example, only the sites 43-45 in the upper strand and 58-61 in the lower strand were substantially more reactive upon looping (Figure 2C, bottom), while the other sites are used poorly, whether or not the DNA spacer is looped. This may be due to sequence bias or to non-uniform bending of the spacer DNA, as suggested by some electron microscopic images (Krämer *et al.*, 1987). The observation that preferred sites for MLV IN and the DNase I are distributed regularly along the looped spacer DNA (Figure 2C) argues, however, against the latter possibility. It is also noteworthy that the preferred site for MLV IN in the upper strand at position 65 has no counterpart on the opposite strand. Presumably this corresponding site on the lower strand is sterically blocked by the *lac* repressor–DNA complex.

To investigate whether the degree of bending correlates with the effect on integration, we tested another construct, *lac*-168-*lac*, with a longer (168 bp) spacer between the two operators. Concomitant with the lower degree of



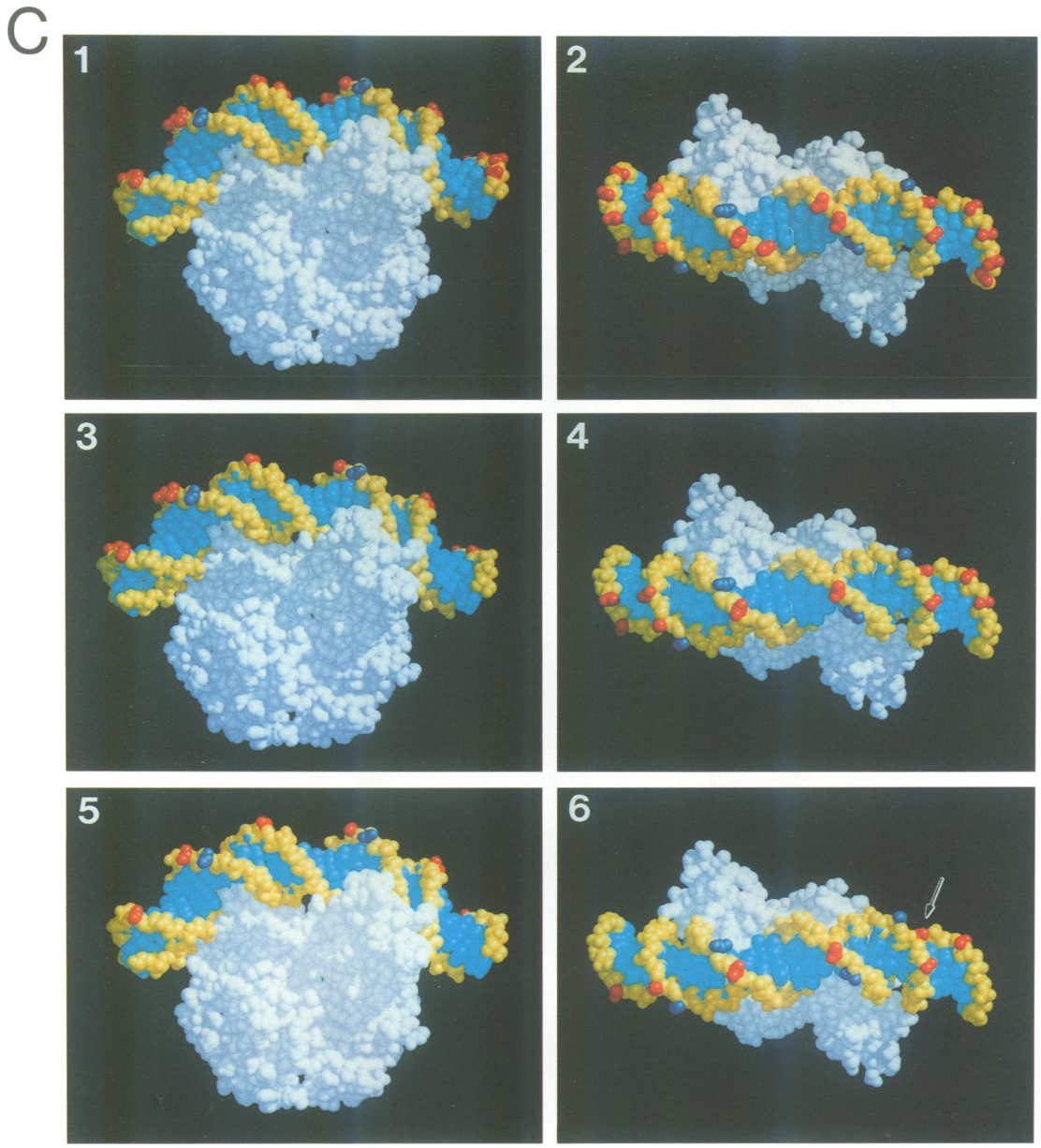


Fig. 1. Integration by MLV IN and HIV IN into the CAP site of the *E.coli lac* operon in the presence or absence of CAP binding. (**A**) PCR analysis of the integration products. At the top the following are indicated: the source of integration activity (HIV IN or MLV IN); the position of the target primer (arrows) relative to the CAP recognition motifs (boxes) used in the PCR analysis; and the presence (+) or absence (−) of cAMP, which induces CAP binding. Target primer BS814+ was used to analyze the upper strand (lanes 2–5) and BS1001- to analyze the lower strand (lanes 6–9). Lane 1 is a sequencing ladder used as a marker. The map positions in the target DNA are indicated to the left of the first and the last two panels. Solid bars to the right of each panel indicate integration footprints. Positions within these footprints at which integration occurs more efficiently on CAP-bound DNA are indicated with black arrowheads numbered with the position in the CAP binding site. The target DNA was prepared from Bluescript KS+ as a 3′-end labeled **Not*I−*Afl*III fragment (label indicated with a star). (B) Distribution of integration sites on CAP-bound DNA. The sequence of the CAP binding site is shown with boxes indicating the major recognition motif. The numbering of the bases is relative to the transcriptional start site (Reznikoff and Abelson, 1980). Black arrowheads indicate integration sites over the region of CAP binding that are used at least as efficiently on CAP-bound as on naked DNA. Integration sites that are more efficiently used than on naked DNA are additionally labeled with the position number. DNase I hypersensitive sites in the CAP−DNA complex, cleaved at least as efficiently as with naked DNA, are indicated with white arrowheads. The extensions of the footprints (black bars) are the same for both HIV and MLV IN. DNase I footprints are indicated as open bars. Uncertainties in the extent of the protection at the end of one footprint are indicated by interrupted bars. The DNase I data correlate well with those described earlier (Schmitz, 1981; Spassky *et al.*, 1984). (**C**) HIV and MLV integration sites in the crystallographic structure model of the CAP−DNA complex (Schultz *et al.*, 1991). The colors indicate the following: CAP protein, white; DNA bases, cyan; DNA backbone, yellow; phosphates attacked by IN, red; phosphates hypersensitive to DNase I, blue (also indicated in Figure 1B with white arrowheads and a star). The CAP−DNA model is shown either from the side (panels 1, 3 and 5) or from the top (panels 2, 4 and 6). The integration sites used at least as efficiently on CAP-bound DNA as on naked DNA are shown in panel 1 and 2 (for HIV IN) and panels 3 and 4 (for MLV IN). The HIV integration site −58 at the lower strand (see Figure 1B) was omitted because it overlaps with the DNase I hypersensitive site. Panels 5 and 6 display the integration sites that are used by both INs more efficiently on CAP-bound DNA than on naked DNA. The site indicated with an arrow is only used more efficiently by MLV IN. Note that the crystallographic structure (Schultz *et al.*, 1991) was solved with an idealized palindromic CAP binding site, whereas this work was carried out using the CAP binding site in the *lac* operon.

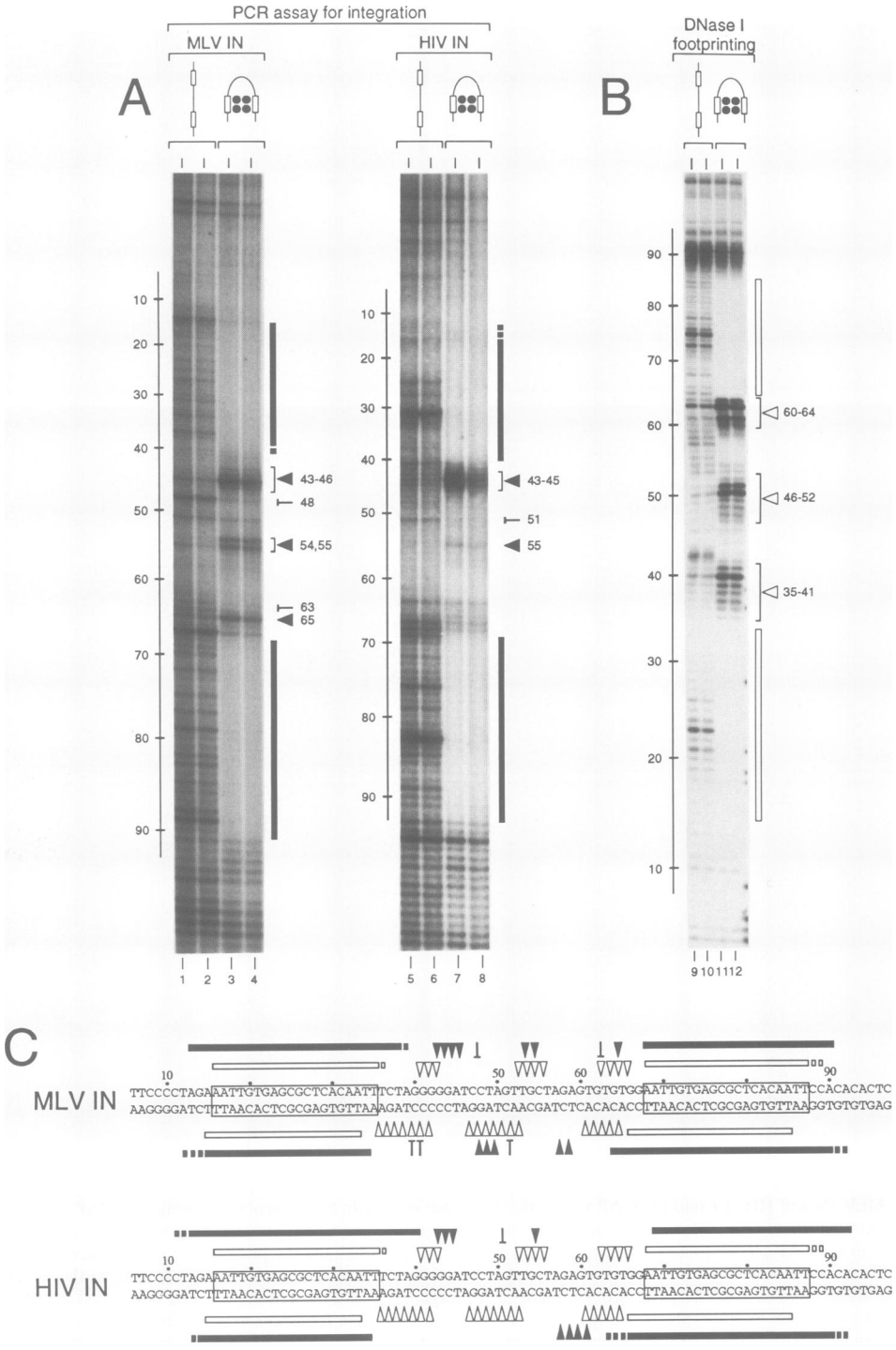


Fig. 2. Integration by MLV IN and HIV IN into a small *lac* repressor-induced loop. The DNA fragment *lac*-52-*lac* was incubated with *lac* repressor protein in approximately stoichiometric amounts producing maximal levels of looped molecules, as determined by a gel shift assay (not shown) or by DNase I footprinting (panel B), and used as integration targets. (**A**) PCR analysis of the integration products. The result obtained with target primer *Bgl*HK4 for the upper strand (see panel C) is shown. Either straight (in the absence of *lac* repressor, lanes 1, 2, 5 and 6) or looped *lac*-52-*lac* DNA (in the presence of *lac* repressor, lanes 3, 4, 7 and 8) were incubated with MLV IN (lanes 1–4) or HIV IN (lanes 5–8). Each integration reaction was performed in duplicate. Integration footprints are indicated with black bars and uncertainties in the extent of the protection with interrupted bars. Black arrowheads indicate sites enhanced for integration upon looping of the DNA. The symbol ⊥ indicates sites reduced for integration upon looping of the DNA. The map positions in the target DNA are indicated to the left of each panel. The *lac*-52-*lac* DNA was prepared from pBKS+52 as a 3′-end labeled **Xho*I−*Bgl*I fragment (label indicated with a star). (**B**) DNase I footprinting of the lower strand of *lac*-52-*lac* DNA in the absence (lanes 9 and 10) or presence (lanes 11 and 12) of *lac* repressor. Each DNase I footprinting reaction was performed in duplicate. Footprints are indicated with open bars. Open arrowheads indicate sites more reactive to DNase I upon looping. Enhancement of cleavage frequency upon looping is maximally 5-fold as measured with a PhosphorImager. The map positions in the target DNA are indicated to the left. The *lac*-52-*lac* DNA was prepared from pBKS+52 as a 3′-end labeled **Not*I−*Xho*I fragment. (**C**) Distribution of integration sites on *lac* repressor-induced loops of *lac*-52-*lac* DNA with MLV IN (top) or HIV IN (bottom). The palindromic *lac* operator sequences are boxed. Symbols are as described in A and B.

bending expected of this longer spacer, the effects on integration sites in the looped region were much less dramatic than with the shorter (52 bp) spacer. Integration site selection by HIV IN is altered subtly but reproducibly upon looping *lac*-168-*lac* (Figure 3A: compare lanes 3 and 4 with lanes 1 and 2), with stimulatory effects of no more than 1.5-fold as measured by phosphorimaging (not shown). The relative distribution of preferred HIV integration sites (Figure 3A) and DNase I cleavage sites (Figure 3B; direct comparison not shown) is similar to that observed with looped *lac*-52-*lac* DNA and CAP–DNA complexes. With MLV IN, however, no change in the integration pattern is observed upon looping *lac*-168-*lac* (Figure 3A, compare lanes 7 and 8 with 5 and 6). The possibility that contaminating nucleases in the IN preparations might nick the target DNA and relieve the distortion in the loop before the integration events occur was eliminated by showing that the DNase I digestion pattern on looped *lac*-168-*lac* DNA is the same before and after preincubation with MLV IN or HIV IN (Figure 3B, compare lanes 17–20 with lanes 9–16).

These experiments suggest that DNA bending can induce integration sites preferred by MLV and HIV IN. Attachment of a protein to the DNA bend, as in CAP–DNA complexes or nucleosomes, is not essential for such stimulatory effects, and the effects are less striking with less severely bent DNA than with more severely bent DNA.

Integration into intrinsically bent A-tract DNA

DNA sequence can sometimes cause DNA to be bent, as most dramatically observed in kinetoplast DNA of *Leishmania tarentolae* (Marini *et al.*, 1982; for review see Crothers *et al.*, 1990). In these sequences, appropriate phasing of so-called A-tracts (homopolymeric dA–dT) leads to bending of the DNA, as indicated by anomalous migration in non-denaturing gels.

We have used as an integration target an artificial A-tract DNA which resembles such naturally occurring sequences (Kahn and Crothers, 1992). The DNA contains six phased A-tract repeats; each repeat consists of six A residues followed by four or five G or C residues, and each is bent by 18–20°. Hence, with six phased repeats the DNA is bent between 108° and 120° over ~65 bp. Integration into this target by MLV IN produces a striking pattern of preferred sites separated by ~10 bp, with essentially no integration occurring between these sites (Figure 4A). This regular distribution of alternating preferred sites and non-utilized sites gradually disappears in the adjacent DNA lacking A-tracts (compare upper and lower portions within each lane). The efficiency of integration at the preferred sites in the A-tract DNA is about the same as at the most preferred sites in the non-A-tract DNA.

Incorporation of these data into a model of A-tract DNA (Crothers *et al.*, 1990) reveals that integration in the A-tract region is favored at sites on the outer face of the bend (Figure 4B), whereas sites on the inner face of the DNA helix are poorly used for integration, consistent with the observation made with the other model targets containing bent DNA. From these experiments we cannot determine unambiguously whether this distinct distribution of integration sites reflects a structural bias or a sequence bias for A-tract DNA, since the structure of A-tract DNA is dictated by the sequence. Only in cases where a given sequence can be manipulated structurally (e.g. by *lac* repressor-induced looping) can the role of structure be clearly demonstrated.

Discussion

Retroviral integration site selection is affected by target DNA bending and associated proteins

The observation that retroviral integration can occur more efficiently into nucleosomal DNA than into nucleosome-free DNA (Pryciak and Varmus, 1992) was unexpected since nucleosomes have repressive effects on several biochemical processes (for recent reviews on chromatin see Wolffe, 1992; Adams and Workman, 1993, and references therein). To understand this phenomenon, we have tested integration by purified MLV IN and HIV IN into various model targets that mimic properties of nucleosomal DNA (Table I).

All of our findings are consistent with the idea that bending of target DNA can promote the use of integration sites on the outer face of the bend. CAP, which bends its recognition site upon binding, still allows the usage of sites of the binding region on the outer, exposed face of the DNA helix and, in some cases, these sites are even more reactive than on the naked DNA (Figure 1). On the other hand, CAP blocks those sites that are located on the inner face of the DNA bend, closer to the CAP protein surface. To allow integration into the protein-bound region, the DNA must likely be bent, since integration is entirely blocked in DNA regions bound by *lac* repressor (Figure 2) or by α2 repressor (Pryciak and Varmus, 1992); neither of these repressor proteins bends their cognate site (Wolberger *et al.*, 1991; Gralla, 1992). As observed with CAP–DNA complexes, sites preferred for integration into *lac* repressor-induced loops and intrinsically bent A-tract DNA are located at the outer face of the DNA bend, whereas sites reduced or blocked for integration are found at the inner face of the bend. Most importantly, experiments with *lac* repressor-induced loops show that stimulation of integration by bending of the target DNA can occur in the absence of histones or other protein surfaces attached to the bent region and that the degree of stimulation of integration upon bending correlates with the degree of bending.

The characteristic pattern of integration sites on nucleosomes can be explained by the observations made in the present work. On the one hand, the creation of new, efficiently used integration sites in nucleosomal DNA seems to be primarily due to bending of the nucleosomal DNA (i.e. the presentation of the DNA in a favorable configuration). In contrast, the blockage of some integration sites in nucleosomal DNA is probably due to steric interference between the nucleosomal core and IN.

Many sites in nucleosomal DNA are stimulated to a degree similar to what is observed with CAP–DNA complexes or *lac* repressor-induced looping of *lac*-52-*lac*, i.e. ~4-fold or less. However, at some sites in nucleosomal targets, integration appears to be enhanced more dramatically (up to 20- to 50-fold; Pryciak and Varmus, 1992). Stimulatory effects in this range were not observed with our models. Possible explanations for this are: (i) the number of potential integration sites in the model targets

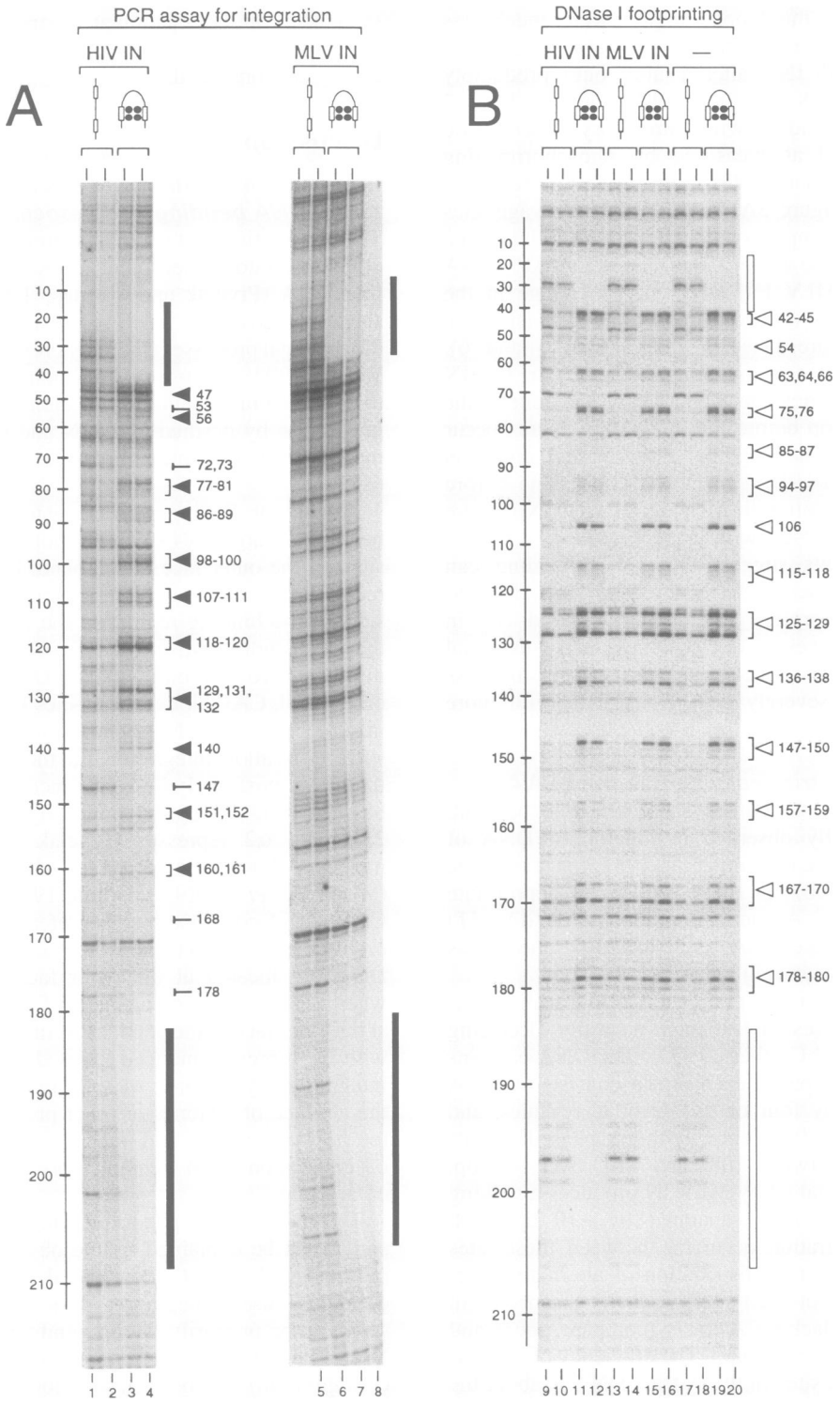


Fig. 3. Integration by MLV IN and HIV IN into a large *lac* repressor-induced loop. The DNA fragment *lac*-168-*lac* complexed with *lac* repressor was prepared as in Figure 2 and used as integration target. (**A**) PCR analysis of the integration products. The result obtained with target primer 168Lop1122− for the upper strand is shown. Either straight (in the absence of *lac* repressor, lanes 1, 2, 5 and 6) or looped *lac*-168-*lac* DNA (in the presence of *lac* repressor, lanes 3, 4, 7 and 8) were incubated with HIV IN (lanes 1–4) or MLV IN (lanes 5–8). Each integration reaction was performed in duplicate. Symbols are as described in Figure 2. The *lac*-168-*lac* DNA was prepared from p168X2BSA1 as a 3′-end labeled **Asp*718−*Nhe*I fragment (label indicated with a star). (**B**) DNase I footprinting of *lac*-168-*lac* DNA in the presence or absence of *lac* repressor. Lanes 17–20 show a typical DNase I footprint experiment as described in Figure 2B. Lanes 9–16 display an experiment to check for potential nucleases in the IN preparations which would reduce the distortion of the DNA. *Lac* repressor-induced loops or the corresponding straight DNA were incubated with HIV (lanes 9–12) or MLV IN (lanes 13–16) as in panel A. Subsequently the reactions were deproteinized, incubated again with *lac* repressor to form loops and analyzed by DNase I footprinting. Each DNase I footprinting reaction was performed in duplicate. Enhancement of cleavage frequency upon looping is maximally 2-fold as measured with a PhosphorImager. Symbols are as described in Figure 2. The fragment used in this experiment was the same as described in panel A.

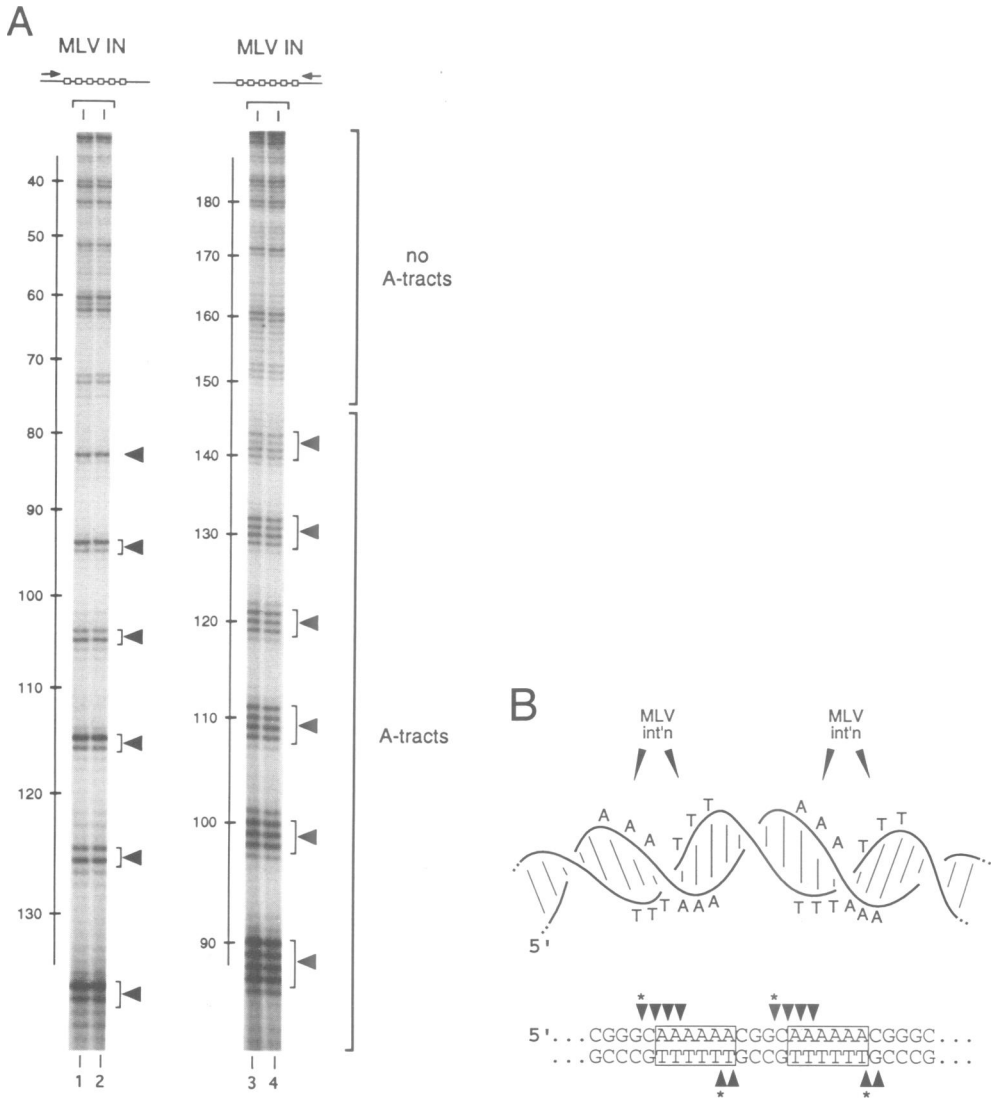


Fig. 4. MLV integration into intrinsically bent A-tract DNA. (**A**) PCR analysis of the integration products. Black arrowheads indicate MLV integration sites in the A-tract DNA that are at least as efficiently used as the most frequent events at sites in the non-A-tract DNA. The left panel displays the result for the lower and the right panel the result for the upper strand. The map positions in the target DNA are indicated to the left of each panel. (**B**) Display of preferred MLV integration sites in a model of A-tract DNA (Crothers *et al.*, 1990). Two consecutive A-tract repeats are indicated as a sequence (bottom) and a structure (top). Black arrowheads marked with a star indicate sites that are not preferred in every A-tract repeat of the target DNA (see for example A-tract repeat at position 83 in the left panel and A-tract repeats at positions 100 and 120 in the right panel of panel A).

used here is much smaller than in the nucleosomal targets used by Pryciak and Varmus (1992). Since only a fraction of preferred sites in nucleosomal DNA is enhanced to the most dramatic level, we may have to search through a larger number of bent sequences to find one that shows this level of reactivity. (ii) DNA bending by CAP widens the exposed minor groove extensively but the adjacent exposed major groove minimally (Schultz *et al.*, 1991). In contrast, both the minor and major grooves are widened in exposed regions of nucleosomal DNA (Richmond *et al.*, 1984). A preference by IN for a widened major groove might enhance use of integration sites less in CAP-bound DNA than in nucleosomal DNA. (iii) DNA in our models may not be bent as sharply as DNA in nucleosomes. The construct *lac*-52-*lac* produces half circles ~50 bp in length, bending DNA ~90° per 25 bp. In contrast, bending of nucleosomal DNA averages ~90° per 20 bp, but is even greater at four positions that are ±1.5 and ±3.5 turns away from the dyad axis (Richmond *et al.*, 1984). Interestingly, very recent work with nucleosomal core particles containing mixed DNA sequences suggests that the most frequent integration events by HIV IN occur at these four positions of sharper-than-average bending (Pruss *et al.*, 1994). Hence, the smaller bend angle produced with *lac* repressor-induced looping might not fully mimic the sharpest of nucleosomal bends. (iv) An interaction between histones and IN may also contribute to the creation of highly preferred sites in nucleosomal DNA.

The relationship we observe between IN reactivity and DNA geometry suggests the possibility that the non-random distribution of integration sites in naked DNA may result from a preference for sequence-dependent variations in DNA structure (e.g. widening of the major or minor grooves). Moreover, the finding that avian leukosis virus integration *in vitro* can be affected by DNA methylation (Kitamura *et al.*, 1992) may reflect a

methylation-dependent change in DNA structure rather than a direct interaction between the integration machinery and methyl groups.

Is DNA bending an intrinsic part of the integration reaction?

Given the stimulatory effect of target DNA bending on integration (or more specifically on the strand transfer reaction), it is feasible that the bending is an intrinsic part of the integration reaction. This seems to be the case for the DNase I cleavage reaction. Crystal structures of DNase I complexed to DNA indicate that the protein interacts exclusively with phosphates and bases in the minor groove (Suck *et al.*, 1988; Lahm and Suck, 1991). These interactions open up the minor groove and bend the DNA away from the protein, suggesting that these distortions are intrinsic features of the DNase I recognition and cleavage reaction. Prebending of the DNA substrate may therefore create sites that resemble intermediates of the DNase I digestion reaction, and thus become preferred cleavage sites. Similarly, our experiments suggest that nucleosomal cores may serve as scaffolds for presenting the DNA in a configuration that reflects an intermediate of the integration reaction. In a reciprocal view, the integration machinery may have evolved to use such an intermediate because most of eukaryotic DNA is bent around nucleosomal cores.

The question of exactly how DNA bending might be involved in the strand transfer reaction remains to be answered. Nevertheless, our observations prompt us to speculate about potential role(s) for DNA bending in integration by drawing analogies with other biochemical reactions. Many biochemical processes, as diverse as restriction endonucleolytic cleavage, recombination, replication, transcription and chromatin assembly, involve DNA bending (for reviews see Echols, 1990; Travers, 1990; Schleif, 1992; van der Vliet and Verrijzer, 1993). Three functions of DNA bending deserve special mentioning in this context.(i) Increase of affinity and specificity of protein–DNA interactions. *Eco*RI, for example, bends the cognate site and increases affinity and cleavage specificity as a consequence of additionally formed hydrogen bonds (McClarin *et al.*, 1986). (ii) Participation in the enzymatic catalysis. *Eco*RV, for example, exclusively bends its cognate site and creates a high affinity site for Mg^{2+}, which is necessary for cleavage (Winkler *et al.*, 1993). (iii) Induction of DNA unpairing. This was suggested on theoretical grounds by Ramstein and Lavery (1988) and most recently experimentally supported with model DNA molecules by Kahn *et al.* (1994).

Target DNA bending could increase either the affinity of IN for the target DNA (K_m) or the reaction rate (k_{cat}). Affinity for the target DNA could be increased by widening the major and/or minor groove(s) on the outside of the bend. Bending could also create a high affinity site for Mn^{2+}, the metal cation essential for the integration reaction. It is also conceivable that the scissile phosphates are brought close to the active site of IN or rendered more reactive upon bending of the target. Another possibility is that the strand transfer reaction requires local denaturation of the target site, which might, in turn, be facilitated by bending. Local unpairing seems to occur in the 3′-processing reaction (B.Scottoline and P.O.Brown, personal communication) which, like the strand transfer reaction, is a transesterification step (Engelman *et al.*, 1991; Vink *et al.*, 1992; reviewed by Mizuuchi, 1992) that could be catalyzed by the same active center of IN (reviewed by Vink and Plasterk, 1993). According to this possibility, mismatches in the target DNA might be hot spots for integration and negatively supercoiled target DNA might be more efficiently used as a target than relaxed or positively supercoiled DNA.

Integration site selection in vitro versus in vivo

Previous studies document a significant bias among potential integration sites *in vivo*, implicating a possible influence of host factors on integration target site selection (for review see Sandmeyer *et al.*, 1990; Craigie, 1992; Bushman, 1993). Transcribed regions and/or DNase I sensitive sites were suggested to be favored (Rohdewohld *et al.*, 1987; Mooslehner *et al.*, 1990), whereas the highly preferred sites found by Shih *et al.* (1988) could not be correlated with any particular aspect of nuclear organization. It is likely that steric accessibility and the degree of target DNA bending will also play an important role in retroviral integration site selection in cells. Consistent with a role of target DNA bending *in vivo*, a possible association between integrated viral DNA and intrinsically bent host DNA has recently been suggested (Milot *et al.*, 1994). To test the effects of steric accessibility and target DNA bending directly, it will be necessary to analyze integration site selection *in vivo* using model targets, such as the ones described here. This could be done by including them as parts of SV40 or BPV episomes, which we have shown to be efficiently used as MLV integration targets *in vivo* (Pryciak *et al.*, 1992; H.-P.Müller and H.E.Varmus, unpublished observations).

Additional properties of nuclear organization will probably need to be taken into account to understand target choice *in vivo* fully. It is possible that specific interactions between host DNA binding proteins and IN direct the integration machinery to certain regions of the host genome, as suggested for the yeast retrotransposon Ty3 (Chalker and Sandmeyer, 1992). Other properties, such as subnuclear localization of viral nucleoprotein complexes or the cell cycle stage during integration (Roe *et al.*, 1993), may impose additional constraints. The influence of features, such as the transcriptional status or the degree of chromatin condensation at given regions of the host genome, on the other hand, might be explained with the principles of steric accessibility and target DNA bending or distortion described here.

Materials and methods

Plasmids and proteins

The construct pBKS+52 (used to prepare *lac*-52-*lac* DNA) is a Bluescript KS+ vector (Stratagene) containing the *Xho*I–*Eco*RI fragment of pHK52 (gift from H.Krämer and B.Müller-Hill; Krämer *et al.*, 1987). The construct p168X2BSA1 (used to prepare *lac*-168-*lac* DNA) is a modified Bluescript KS+ vector containing the *Hae*III–*Hin*dIII fragment of pKO6168 (gift from H.Krämer and B.Müller-Hill; Krämer *et al.*, 1987).

CAP protein was a gift from C.Scafe and T.A.Steitz. *lac* repressor protein was purchased from Stratagene. HIV IN, a gift from A.D.Leavitt, was produced in yeast and prepared as described (Leavitt *et al.*, 1992). MLV IN, a gift from L.Shiue, was prepared from yeast similarly to HIV IN, with the exception that the Affigel–heparin and phenyl Sepharose columns were replaced by a 30% ammonium precipitation.

Integration into CAP–DNA complexes

In the experiments with CAP–DNA complexes the *AflIII–Asp718 (397 bp), *Asp718–AflIII (397 bp) and *NotI–AflIII (483 bp) fragments were retrieved from Bluescript KS+, 3′-end labeled with ^{32}P using Klenow fragment (label indicated with asterisk) and purified on native polyacrylamide gels. The CAP binding reaction was performed for 15 min at 25°C in a total volume of 150 µl with ~6 ng DNA and 150 ng CAP in 10 mM Tris–HCl (pH 8), 20 mM KCl, 1 mM DTT, 10 mM MgAc, 0.1 mM EDTA, 100 µg/ml BSA. CAP was present in all reactions and binding was induced by adding 200 µM cAMP (Sigma). For the gel shift assays, 15 µl of the binding reaction were loaded on a native 4% polyacrylamide gel prepared and subjected to electrophoresis as described by Krämer *et al.* (1987), except that 100 µM cAMP were added to the running buffer. The DNase I footprinting reaction was performed with 150 µl binding reaction containing 10 µg/ml sonicated salmon sperm DNA and 0.55 U RQ1 DNase I (Promega). After incubation for 1 min at 25°C the reaction was stopped with proteinase K and separated on a sequencing gel. For the integration reactions 15 µl integration buffer (500 mM MOPS pH 7, 150 mM $MnCl_2$, 100 mM DTT, 1 mg/ml BSA), 1 µl 2.25 µM processed *att* site oligos (annealed U5MLVA27 and U5MLVB29 for MLV IN or annealed HIVU527+ and HIVU529– for HIV IN) and ~0.15 µg MLV IN or 0.1 µg HIV IN were added to 135 µl binding reaction, resulting in a total reaction volume of 155 µl. After 5 min at 25°C the reactions were stopped with proteinase K and prepared for PCR as described (Pryciak and Varmus, 1992). The recombinants were resuspended in a final volume of 50 µl 10 mM Tris–HCl pH 7.5, 1 mM EDTA (1 µl was usually used per PCR). The PCR assays for integration site distribution in this and the subsequent experiments were performed as described (Pryciak and Varmus, 1992). The mapping of integration sites was also performed as described by Pryciak and Varmus (1992), using amplified integration recombinants and dideoxy sequencing ladders as markers. DNA fragments obtained with dideoxy sequencing migrate ~0.5–1 nucleotide faster than the deoxynucleotide- terminated fragments of the same size obtained with PCR. BS814+ and BS1001– served as target primers. The viral primers were U5MLVA27 for MLV integration recombinants and HIVU527+ for HIV integration recombinants. Molecular graphics images were produced using the MidasPlus program (Ferrin *et al.*, 1988; Huang *et al.*, 1991) from the Computer Graphics Laboratory, UCSF.

Integration into lac repressor-induced loops

In the experiments with *lac* repressor-induced loops the *Asp718–NheI (489 bp) and *XhoI–Asp718 (280 bp) as well as *NotI–XhoI (303 bp), *HindII–BglI (338 bp) and *XhoI–BglI (503 bp) fragments were retrieved from p168X2BSA1 (for *lac*-168-*lac*) and pBKS+52 (for *lac*-52-*lac*), respectively. The fragments were 3′-end labeled with ^{32}P by filling in with Klenow fragment (sites indicated with asterisk) and purified on polyacrylamide gels. The binding reaction was performed for 20 min at 25°C in a total volume of 150 µl with ~0.2 ng fragment and 0.5 ng *lac* repressor in 10 mM Tris–HCl pH 8, 10 mM KCl, 0.1 mM DTT, 10 mM MgAc, 0.1 mM EDTA, 100 µg/ml BSA. The exact amount of *lac* repressor required to produce a maximal amount of looped molecules was determined by gel shift analysis and DNase I footprinting experiments as described (Krämer *et al.*, 1987). The integration experiments and PCR assays for integration site distribution were performed as described above. The target primers in the PCR assay were BS631+ and *Bgl*HK4 for *lac*-52-*lac*, 168Lop827+ and Ovec-12/-32 for *lac*-168-*lac*.

Integration into A-tract DNA

The A-tract DNA 9A17 was produced by PCR using primers 2 and 1B and *Bst*NI-digested plasmid 11A17 (gift from J.D.Kahn and D.M.Crothers) as template and the PCR product was digested with *Cla*I and circularized with T4 ligase as described (Kahn and Crothers, 1992). The targets for integration 9A17*Nla* and 9A17*Ban* were produced by linearization with *Nla*III or *Ban*I, respectively. Integration reactions were carried out in a total volume of 10 µl with 1 ng target DNA (9A17*Nla* or 9A17*Ban*) and ~0.4 µg MLV IN in 50 mM MOPS pH 7, 15 mM $MnCl_2$, 10 mM DTT, 100 µg/ml BSA. After 5 min at 25°C the reaction was stopped with proteinase K and prepared for PCR as described above. The target primers MC4 and MC3 were used with targets 9A17*Nla* and 9A17*Ban*, respectively.

Oligonucleotides

The following oligonucleotides (prepared by the Biomolecular Resource Center, UCSF) were used in this study: BS814+: GAGCTTGGCGTAATCATGGTCATAGC; BS1001: GCCCAATACGCAAACCGCCTCTCC; BS631+: CGGCCAGTGAATTGTAATACGACTC; *Bgl*HK4: GAAGATCTGCTTCCGGCTCGTATGTTGTGTGG; 168Lop827+: GCAGCGAGTCAGTGAGCGAGGAAG; Ovec-12/-32: GCAGTGCTCTGCCTTTTATGC; MC3: CCTAGGTCTAGAATTCAGCTGT; MC4: CCATGGAATCGATGAATTCACG; U5MLVA27: TGACTACCCGTCAGCGGGGGTCTTTCA; U5MLVB29: AATGAAAGACCCCCGCTGACGGGTGTCA; HIVU527+: TTTAGTCAGTGTGGAAAATCTCTAGCA; HIVU529-: ACTGCTAGAGATTTTCCACACTGACTAAA.

Acknowledgements

We are grateful to the following people for generous gifts of reagents: Helmut Krämer and Benno Müller-Hill for the plasmids pHK52 and pKO6168; Jason Kahn and Don Crothers for the plasmid 11A17; Charles Scafe and Tom Steitz for CAP protein; Lily Shiue for MLV IN protein; and Andy Leavitt for HIV IN protein. We also thank Andy Leavitt, Susanna Lee and Peter Pryciak for discussions throughout this work and Andy Leavitt, Susanna Lee, Michael Müller-Immergluck, John Murphy and Peter Pryciak for comments on the manuscript. This work was supported by an NIH grant to H.E.V., an NIH grant to the Computer Graphics Laboratory (USCF), and by a fellowship from the Swiss National Science Foundation to H.-P.M. H.E.V. was an American Cancer Society Research Professor.

References

Adams,C.C. and Workman,J.L. (1993) *Cell*, **72**, 305–308.
Bowerman,B., Brown,P.O., Bishop,J.M. and Varmus,H.E. (1989) *Genes Dev.*, **3**, 469–478.
Brown,P.O., Bowerman,B., Varmus,H.E. and Bishop,J.M. (1989) *Proc. Natl Acad. Sci. USA*, **86**, 2525–2529.
Bushman,F.D. (1993) *Curr. Biol.*, **3**, 533–535.
Bushman,F.D., Fujiwara,T. and Craigie,R. (1990) *Science*, **249**, 1555–1558.
Chalker,D.L. and Sandmeyer,S.B. (1992) *Genes Dev.*, **6**, 117–128.
Coffin,J.M. (1990) In Fields,B.N. and Knipe,D.M. (eds), *Virology*. Raven Press, New York, pp. 1437–1500.
Craigie,R. (1992) *Trends Genet.*, **8**, 187–190.
Craigie,R., Fujiwara,T. and Bushman,F. (1990) *Cell*, **62**, 829–837.
Crothers,D.M., Haran,T.E. and Nadeau,J.G. (1990) *J. Biol. Chem.*, **265**, 7093–7096.
Crothers,D.M. and Steitz,T.A. (1992) In McKnight,S.L. and Yamamoto,K.R. (eds), *Transcriptional Regulation*. Cold Spring Harbor Laboratory Press, Plainview, NY, pp. 501–534.
Echols,H. (1990) *J. Biol. Chem.*, **265**, 14697–14700.
Engelman,A., Mizuuchi,K. and Craigie,R. (1991) *Cell*, **67**, 1211–1221.
Ferrin,T.E., Huang,C.C., Jarvis,L.E. and Langridge,R. (1988) *J. Mol. Graphics*, **6**, 13–27.
Fitzgerald,M.L., Vora,A.C., Zeh,W.G. and Grandgenett,D.P. (1992) *J. Virol.*, **66**, 6257–6263.
Fujiwara,T. and Mizuuchi,K. (1988) *Cell*, **54**, 497–504.
Goff,S.P. (1992) *Annu. Rev. Genet.*, **26**, 527–544.
Gralla,J.D. (1992) In McKnight,S.L. and Yamamoto,K.R. (eds), *Transcriptional Regulation*. Cold Spring Harbor Laboratory Press, Plainview, NY, pp. 629–642.
Huang,C.C., Pettersen,E.F., Klein,T.E., Ferrin,T.E. and Langridge,R. (1991) *J. Mol. Graphics*, **9**, 230–236.
Kahn,J.D. and Crothers,D.M. (1992) *Proc. Natl Acad. Sci. USA*, **89**, 6343–6347.
Kahn,J.D., Yun,E. and Crothers,D.M. (1994) *Nature*, **368**, 163–166.
Katz,R.A., Merkel,G., Kulkosky,J., Leis,J. and Skalka,A.M. (1990) *Cell*, **63**, 87–95.
Kitamura,Y., Lee,Y.M.H. and Coffin,J.M. (1992) *Proc. Natl Acad. Sci. USA*, **89**, 5532–5536.
Krämer,H., Niemöller,M., Amouyal,M., Revet,B., von Wilcken-Bergmann,B. and Müller-Hill,B. (1987) *EMBO J.*, **6**, 1481–1491.
Lahm,A. and Suck,D. (1991) *J. Mol. Biol.*, **221**, 645–667.
Leavitt,A.D., Rose,R.B. and Varmus,H.E. (1992) *J. Virol.*, **66**, 2359–2368.
Marini,J.C., Levene,S.D., Crothers,D.M. and Englund,P.T. (1982) *Proc. Natl Acad. Sci. USA*, **79**, 7664–7668.
McClarin,J.A., Frederick,C.A., Wang,B.C., Greene,P.J., Boyer,H.W., Grable,J.C. and Rosenberg,J.M. (1986) *Science*, **234**, 1526–1541.
Milot,E., Belmaaza,A., Rassart,E. and Chartrand,P. (1994) *Virology*, **201**, 408–412.
Mizuuchi,K. (1992) *J. Biol. Chem.*, **267**, 21273–21276.
Mooslehner,K., Karls,U. and Harbers,K. (1990) *J. Virol.*, **64**, 3056–3058.

Müller,H.-P., Pryciak,P.M. and Varmus,H.E. (1993) *Cold Spring Harbor Symp. Quant. Biol.*, **58**, 533–541.
Pruss,D., Bushman,F.D. and Wolffe,A.P. (1994) *Proc. Natl Acad. Sci. USA*, **91**, 5913–5917.
Pryciak,P.M. and Varmus,H.E. (1992) *Cell*, **69**, 769–780.
Pryciak,P.M., Müller,H.-P. and Varmus,H.E. (1992) *Proc. Natl Acad. Sci. USA*, **89,** 9237–9241.
Ramstein,J. and Lavery,R. (1988) *Proc. Natl Acad. Sci. USA*, **85**, 7231–7235.
Reznikoff,W.S. and Abelson,J.N. (1980) In Miller,J.H. and Reznikoff,W.S. (eds), *The Operon.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 221–243.
Richmond,T.J., Finch,J.T., Rushton,B., Rhodes,D. and Klug,A. (1984) *Nature*, **311**, 532–537.
Roe,T.Y., Reynolds,T.C., Yu,G. and Brown,P.O. (1993) *EMBO J.*, **12**, 2099–2108.
Rohdewohld,H., Weiher,H., Reik,W., Jaenisch,R. and Breindl,M. (1987) *J. Virol.*, **61**, 336–343.
Sandmeyer,S.B., Hansen,L.J. and Chalker,D.L. (1990) *Annu. Rev. Genet.*, **24**, 491–518.
Schleif,R. (1992) *Annu. Rev. Biochem.*, **61**, 199–223.
Schmitz,A. (1981) *Nucleic Acids Res.*, **9**, 277–292.
Schultz,S., Shields,G.C. and Steitz,T.A. (1991) *Science*, **253**, 1001–1007.
Shih,C.-C., Stoye,J.P. and Coffin,J.M. (1988) *Cell*, **53,** 531–537.
Spassky,A., Busby,S. and Buc,H. (1984) *EMBO J.*, **3**, 43–50.
Suck,D., Lahm,A. and Oefner,C. (1988) *Nature*, **332**, 465–468.
Travers,A.A. (1990) *Cell*, **60**, 177–180.
van der Vliet,P.C. and Verrijzer,C.P. (1993) *BioEssays*, **15**, 2532.
Varmus,H.E. and Brown,P. (1989) In Berg,D.E. and Howe,M.M. (eds), *Mobile DNA.* American Society of Microbiology, Washington DC, pp. 53–108.
Vink,C. and Plasterk,R.H.A. (1993) *Trends Genet.*, **9**, 433–437.
Vink,C., Yeheskiely,E., van der Marel,G.A., van Boom,J.H. and Plasterk,R.H.A. (1992) *Nucleic Acids Res.*, **19**, 6691–6698.
Whitcomb,J.M. and Hughes,S.H. (1992) *Annu. Rev. Cell Biol.*, **8**, 275–306.
Winkler,F.K., Banner,D.W., Oefner,C., Tsernoglou,D., Brown,R.S., Heathman,S.P., Bryan,R.K., Martin,P.D., Petratos,K. and Wilson,K.S. (1993) *EMBO J.*, **12**, 1781–1795.
Wolberger,C., Vershon,A.K., Liu,B., Johnson,A.D. and Pabo,C.O. (1991) *Cell*, **67**, 517–528.
Wolffe, A.P. (1992) *Chromatin: Structure and Function.* Academic Press, San Diego, CA.

Received on June 20, 1994; revised on July 26, 1994